

Advanced Placement Grading: Lesson Learned

Chris Hakala

Western New England College

Rob McEntarffer

Lincoln Southeast High School

(This essay originally appeared as the monthly “E-xcellence in Teaching” e-column in the *PsychTeacher Electronic Discussion List* for December 2004.)

Teachers of psychology often complain that it is very difficult to grade essays effectively and consistently (e.g., Nodine, 1999; Zlokovich, 2004). Although graduate student teachers often spend considerable time discussing effective classroom techniques, they typically spend less time learning to grade effectively. Moreover, although future teachers often learn that there is more than one way to grade essays and that each person should determine what works best for him or her, we have found that applying this knowledge is sometimes difficult. When we began grading exams for Advanced Placement (AP) psychology, it became clear that the method by which we normally graded essays may not have been the most efficient or effective. In this essay, we briefly describe AP grading, our experiences as AP readers, and the impact AP grading has had on our ability to evaluate our students’ work.

AP Grading

Students who have completed AP courses in high school take their AP exams in May. All AP exams follow the same general format, with the AP Psychology exam consisting of 100 multiple-choice and 2 essay questions. The College Board brings together a group of high school and college psychology teachers to hand-score the essays. Although the AP reading has been held at various sites around the United States, it has been held in Daytona Beach, Florida, for the past 4 years.

A leadership team consisting of approximately 20 members arrives early and prepares a scoring system, or rubric. Rubric development is a difficult process. The team starts with a generic scoring guide provided by the test development committee. Next, the leadership team uses textbooks, their experiences, and sample student responses to craft a scoring guide that is easy to understand and apply.

When the readers arrive, they receive 1 day of training with the rubric, which teaches them to grade essays consistently, a very important task given that, in 2004, readers graded over 72,000 booklets, each containing two essays. Upon completion of the training, readers then score essays from 8:30 a.m. until 4:45 p.m. each day. During this time, they are constantly monitored by table leaders who check to make sure they are scoring according to the rubric. Although grading may sound as though it would be difficult, we have found it to be one of the most rewarding experiences of our careers.

Our Experiences

Chris Hakala. My first exposure to grading essays was as a graduate student at the University of New Hampshire (UNH). I was enrolled in a teaching seminar taught by Victor Benassi, winner of the 2003 American Psychological Foundation Award for Teaching Excellence. The graduate program at UNH is well known for preparing its students for careers in academia, and I was thrilled to learn the “tricks of the trade.” During the seminar, we learned about reliability and validity as it applied to testing and assessment, and Victor described techniques that would assist in teaching and grading. For example, we learned that student essays should be scored blindly to avoid bias, and that, as faculty, we need to ensure that all students are treated equitably. After graduation, I took this knowledge with me and believed that I was well versed in the ways of student assessment. After all, I had read McKeachie’s (2002) *Teaching Tips* as well as work by Lowman (1995) and others.

In 1996, I attended the Northeastern Conference of Teachers of Psychology hosted by Barney Beins at Ithaca College. At the time, I was interested in high school psychology and had recently conducted a study demonstrating that college students who took high school psychology fared no better in an introductory psychology course than those who had not. After presenting these data at the conference, I was met by a small but angry group of talented high school teachers who questioned my results. Fortunately, I was also approached by Jane Halonen, who, after seeing my interest in high school psychology, told me about AP grading. After learning more about it, I applied to be a reader.

My first reading was in 1997, and it was quite an experience. To begin with, I arrived late at Clemson University, where the reading was held. Then, at my first training session, I was paired with Dr. Charles Brewer. Needless to say, I was scared, intimidated, and convinced that this would be my first—and last—AP reading.

Ultimately, however, it turned out to be a great experience. Although I learned much during that first reading, the most important thing I learned was the strategy that the College Board and Educational Testing Service (ETS) use to carry out the scoring of a large number of essays. In my years of teaching, I have often struggled with the idea that I was somehow grading essays inefficiently. Further, I often found my mind drifting while I graded them. That is, I graded essays differently as the pile got smaller. I was amazed at the organization of the AP reading and the way the College Board and ETS worked to ensure reliability and validity. The goal, of course, was to provide students with a fair, unbiased reading of their essays, and the entire grading process was designed with that goal in mind.

The process of using a rubric to score the essays along with continuous reliability checks was useful and gave me confidence that I was grading my essays effectively. Over the years, I have tried, with mixed success, to adapt this process to grading essays in my courses. Although using a rubric is wonderfully helpful and I have found that it makes grading essays much easier, adapting the other parts of the reading (e.g., conducting reliability checks) is more difficult. The reading, however, has taught me that scoring can be made much more objective if I attempt to adapt these processes the best I can.

Rob McEntarffer. My experience mirrors Chris's in some ways, but I have viewed it from a different perspective: as a high school teacher rather than a college professor. In my teacher-training program, a very small amount of our coursework covered the topic of assessment. In fact, most of what I learned about classroom assessment was from other high school teachers. I was never taught how to write clear exam items, examine the reliability or validity of those items, or grade student essays.

Then, in 1996, I was invited to be a "rescue reader" at that year's AP grading. One of the other AP readers could not attend, so I was flown in on the second day of the reading and arrived late at night. The next day, I was late for the opening meeting. As I walked in front of the other readers, I heard a kind, southern voice (Dr. Charles Brewer) say "Oh, I see they're inviting high school students now." Although I quickly became friends with Charles and the other readers that year, it was nevertheless a new and somewhat intimidating world to me. Up to that point, I had never thought about the issues integral to AP grading, including how to assess students' knowledge fairly and quickly from their essay responses.

That year, I learned what a clear and complete rubric looks like and its importance for accurate essay grading. Before we started grading, we discussed in detail what to include in the rubric. This process forced us to identify the essential elements that would demonstrate knowledge of a psychological concept. For example, if students said that spontaneous recovery occurs after a response becomes extinct, we discussed whether this answer was sufficient, or whether it was necessary for them to state that a time period must elapse after extinction before the behavior reoccurs. I now construct my own grading rubrics by asking: What essential elements of the concepts do I expect my students to know?

I also learned how important reliability and validity checks are for consistent grading. I was, and continue to be, impressed with the reliability checks built into the AP grading process. I appreciate how each student's essay is considered carefully and how reliability data help ensure that each reader is held to the same standard of grading. Although circumstances prevent me from applying each of these standards to my own grading, I make sure I grade each essay anonymously, and I re-grade some essays to ensure that I am consistently applying the rubric.

Finally, I learned how careful training and hard work on a grading rubric can actually minimize the time spent grading. At first, the thought of grading a huge stack of essays in a week was daunting. Now that I have "become one with the rubric" (the mantra of AP readers), I am in a position to grade essays more effectively and efficiently.

In Conclusion

AP reading has been a terrific experience for both of us. We believe we are better able to score essays and provide students with fair and unbiased feedback on their performance and understanding of material. Teacher training programs would do well to consider the work of AP reading as a model for helping future teachers develop the skills necessary to evaluate material fairly and consistently.

References

- Lowman, J. (1995). *Mastering the techniques of teaching* (2nd ed.). San Francisco: Jossey-Bass.
- McKeachie, W. (2002). *Teaching tips: Strategies, research, and theory for college and university teachers* (11th ed.). Boston: Houghton Mifflin.
- Nodine, B. F. (1999). Why not make writing assignments? In B. Perlman, L. I. McCann, & S. H. McFadden (Eds.). *Lessons learned: Practical advice for the teaching of psychology* (pp. 167-172). Washington, DC: American Psychological Society.
- Zlokovich, M. S. (2004). Grading for optimal student learning. In B. Perlman, L. I. McCann, & S. H. McFadden (Eds.). *Lessons learned: Practical advice for the teaching of psychology, Vol. 2* (pp. 255-264). Washington, DC: American Psychological Society.